US009665633B2

(12) **United States Patent**
Dageville et al.

(10) **Patent No.:** US 9,665,633 B2
(45) **Date of Patent:** May 30, 2017

(57) **ABSTRACT**

Example data management systems and methods are described. In one implementation, a method identifies multiple files to process based on a received query and identifies multiple execution nodes available to process the multiple files. The method initially creates multiple scansets, each including a portion of the multiple files, and assigns each scanset to one of the execution nodes based on a file assignment model. The multiple scansets are processed by the multiple execution nodes. If the method determines that a particular execution node has finished processing all files in its assigned scanset, an unprocessed file is reassigned from another execution node to the particular execution node.

24 Claims, 14 Drawing Sheets